

4-2014

A Meta Analysis of Pancreatic Microarray Datasets Yields New Targets as Cancer Genes and Biomarkers

Nalin C. W. Goonesekere
University of Northern Iowa

Xiaosheng Wang

University of Nebraska Medical Center
[Let us know how access to this document benefits you](#)

See next page for additional authors

Copyright ©2014 Nalin C. W. Goonesekere, Xiaoshen Wang, Lindsey Ludwig, and Chittibabu Guda. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.



This work is licensed under a [Creative Commons Attribution 3.0 License](#).

Follow this and additional works at: https://scholarworks.uni.edu/che_facpub

 Part of the [Chemistry Commons](#)

Recommended Citation

Goonesekere, Nalin C. W.; Wang, Xiaosheng; Ludwig, Lindsey; and Guda, Chittibabu, "A Meta Analysis of Pancreatic Microarray Datasets Yields New Targets as Cancer Genes and Biomarkers" (2014). *Faculty Publications*. 2.

https://scholarworks.uni.edu/che_facpub/2

This Article is brought to you for free and open access by the Faculty Work at UNI ScholarWorks. It has been accepted for inclusion in Faculty Publications by an authorized administrator of UNI ScholarWorks. For more information, please contact scholarworks@uni.edu.

Offensive Materials Statement: Materials located in UNI ScholarWorks come from a broad range of sources and time periods. Some of these materials may contain offensive stereotypes, ideas, visuals, or language.

Authors

Nalin C. W. Goonesekere, Xiaosheng Wang, Lindsey Ludwig, and Chittibabu Guda



A Meta Analysis of Pancreatic Microarray Datasets Yields New Targets as Cancer Genes and Biomarkers

Nalin C. W. Goonesekere^{1*}, Xiaosheng Wang², Lindsey Ludwig¹, Chittibabu Guda^{2,3}

1 Department of Chemistry and Biochemistry, University of Northern Iowa, Cedar Falls, Iowa, United States of America, **2** Department of Genetics, Cell Biology and Anatomy, University of Nebraska Medical Center, Omaha, Nebraska, United States of America, **3** Bioinformatics and Systems Biology Core, University of Nebraska Medical Center, Omaha, Nebraska, United States of America

Abstract

The lack of specific symptoms at early tumor stages, together with a high biological aggressiveness of the tumor contribute to the high mortality rate for pancreatic cancer (PC), which has a five year survival rate of less than 5%. Improved screening for earlier diagnosis, through the detection of diagnostic and prognostic biomarkers provides the best hope of increasing the rate of curatively resectable carcinomas. Though many serum markers have been reported to be elevated in patients with PC, so far, most of these markers have not been implemented into clinical routine due to low sensitivity or specificity. In this study, we have identified genes that are significantly upregulated in PC, through a meta-analysis of large number of microarray datasets. We demonstrate that the biological functions ascribed to these genes are clearly associated with PC and metastasis, and that that these genes exhibit a strong link to pathways involved with inflammation and the immune response. This investigation has yielded new targets for cancer genes, and potential biomarkers for pancreatic cancer. The candidate list of cancer genes includes protein kinase genes, new members of gene families currently associated with PC, as well as genes not previously linked to PC. In this study, we are also able to move towards developing a signature for hypomethylated genes, which could be useful for early detection of PC. We also show that the significantly upregulated 800+ genes in our analysis can serve as an enriched pool for tissue and serum protein biomarkers in pancreatic cancer.

Citation: Goonesekere NCW, Wang X, Ludwig L, Guda C (2014) A Meta Analysis of Pancreatic Microarray Datasets Yields New Targets as Cancer Genes and Biomarkers. PLoS ONE 9(4): e93046. doi:10.1371/journal.pone.0093046

Editor: Jörg D. Hoheisel, Deutsches Krebsforschungszentrum, Germany

Received: June 6, 2013; **Accepted:** February 28, 2014; **Published:** April 16, 2014

Copyright: © 2014 Goonesekere et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The authors have no support or funding to report.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: nalin.goonesekere@uni.edu

Introduction

Pancreatic cancer (PC) is a highly lethal malignancy, and patients with PC have a 5-year survival rate of less than 5% [1]. Thus, even though the incidence of breast cancer is estimated to be 5 times greater than PC, the annual death rates are comparable [2]. In nearly 95% of PC patients there is neither an associated family history of PC nor of diseases known to be associated with an increased risk of PC [3]. The lack of specific symptoms at early tumor stages, together with a high biological aggressiveness of the tumor and resistance to cytotoxic drugs all contribute to the high mortality rate of PC.

This study has been motivated by two reasons. The First is to contribute to the understanding of the fundamental disease etiology of PC by identifying novel candidate cancer genes in pancreatic cancer. The mutations found in a cancer cell genome have generally accumulated over the lifetime of the cancer patient and usually number between 1,000–10,000 [4]. For PC, exome sequencing has revealed that the average number of mutations in exons is about 60 [5]. Driver mutations [6] confer growth advantage on the cells carrying them and are positively selected for during the evolution of a cancer. It has been suggested that common adult epithelial cancers require the activation of between 5–20 such driver genes [7,8]. The identification of driver mutations and the cancer genes that they alter has been a central aim of cancer research; so far, about 500 (2%) of the 22,000 protein-coding genes in the human genome are reported to show

recurrent mutations in cancer with strong evidence that these contribute to cancer development [9] (<http://www.sanger.ac.uk/genetics/CGP/Census/>). However, studies in mice have suggested that more than 2,000 genes, when appropriately altered, may have the potential to contribute to cancer development [10] indicating that the search for cancer genes is far from over. A comprehensive treatment protocol for pancreatic cancer would require first, the identification of all the cancer genes, and next, the ability to modulate the function of these genes through therapeutic intervention. In recent years, the proteins altered by driver mutations have become targets for successful anticancer drug development [11–13].

The second impetus for this study comes from the paucity of biomarkers in PC [14,15]. Improved screening for earlier diagnosis, through the detection of diagnostic and prognostic biomarkers, provides the best hope of increasing the rate of curatively resectable carcinomas. For example, analysis of sequence data has suggested that the time frame from the initiation of the pancreatic tumor to the development of metastatic subclones could be more than ten years [16]. Though many serum markers has been reported to be elevated in patients with pancreatic cancer, so far, most of these markers have not been implemented into clinical routine due to low sensitivity or specificity [14] with the exception of CA 19-9 [17].

Meta-analysis of microarray datasets consists of using statistical techniques to combine results from several studies in order to

increase statistical power and generalizability compared with any single study [18]. This addresses, to some extent, the issues of biological and technical variations, which can have a significant effect on microarray measurements [19]. The previous meta-analysis of microarray datasets on PC was conducted nearly a decade ago by Grutzman and colleagues [20], and the analysis was limited to a few thousand genes.

In this study, we examined the differential gene expression patterns that are replicated across datasets, to create a ranked list of genes overexpressed in PC. We focused our attention only on genes that are overexpressed, since about 80% of cancer genes are dominant acting [4] through either overexpression or constitutive activation of gene product. In this study, we have detected hundreds of genes that were significantly upregulated in pancreatic cancer. The list of overexpressed genes include genes that have not been previously associated with PC as well as new members of gene families that have been associated with PC. We have also identified tens of kinase-encoding genes overexpressed in pancreatic cancer, which are potential therapeutic targets for PC. In this study, we are also able to move towards developing a signature for hypomethylated genes, which could be useful for early detection of PC. We also find that about a third of the putative protein serum biomarkers thus far identified for PC are, in fact, significantly overexpressed in our analysis, indicating that our results could serve as a resource for further experimental studies, in the quest for effective biomarkers for PC.

Materials and Methods

Pancreatic cancer microarray datasets

Nine pancreatic cancer datasets in the Oncomine database [21] that contained a differential analysis of pancreatic cancer vs. normal samples, were included this study (Table 1). Oncomine [21] is the most comprehensive cancer-specific database, currently containing 628 datasets investigating 35 tumor types (Oncomine 4.4 Research Edition). The advantage of using datasets from Oncomine is that prior to inclusion in Oncomine, the microarray datasets (obtained from public resources such as Stanford Microarray Database and the NCBI Gene Expression Omnibus or literature sources) are reviewed by a panel of experts to ensure that they meet certain quality standards [22].

Initial screening of microarray datasets

Prior to combining microarray datasets from different sources, a further quality check was performed on the datasets using the program Venn Mapper [23]. Venn mapper can identify significant similarities between heterologous microarray datasets, by comparing the overlap of differentially expressed genes and calculating a statistical significance using z-values. Briefly, a 2-fold cutoff is used to determine the upregulated genes in a microarray dataset. A list of upregulated genes is established for each microarray, and all pair-wise (except self comparisons) combinations of lists are compared for matching gene-identity (i.e. HUGO gene names). The number of genes commonly upregulated, $R_{observed}$, in any two experiments is determined, and a z-value is calculated to determine whether this number is statistically significant. For two microarrays A and B, the z-value is calculated as follows:

$$z = \frac{R_{observed} - R_{expected}}{\sigma} = \frac{R_{observed} - n_B P_A}{\sqrt{n_B P_A (1 - P_A)}}$$

R = number of genes upregulated in both A and B

n_B = Total number of genes upregulated in B

P_A = Probability of a gene being upregulated in A

Microarrays were clustered based on z-value profiles, and any outliers were identified, and omitted from further analysis. An absolute z-value of >1.96 is equivalent to a p-value of <0.05 .

Obtaining ranked lists of upregulated genes

To identify differentially expressed genes across multiple datasets, we employed a non-parametric 'rank product' method implemented in the RankProd package [24,25]. RankProd is a statistically rigorous but biologically intuitive algorithm, which has been shown to be robust against noise in microarray data [26,27]. RankProd has been shown to have higher sensitivity and specificity compared to other types of meta-analytic tools for microarrays [28]. A list of upregulated genes are selected based on a conservative estimation of the percentage of false positive predictions (pfp), which is also known as the false discovery rate. As recommended, a pfp value of <0.15 [25] was used to set the threshold for genes that are significantly upregulated.

Results and Discussion

Congruency between microarray datasets

The program Venn Mapper [23] was used to perform an initial screening, to determine any broad inconsistencies that exist between the microarray datasets. Analysis was carried out on nine different datasets, and all-to-all pairwise z-values are given in Table 2. Two outliers were identified by this method, namely, Buchholz Pancreas (Pancreatic Ductal Adenocarcinoma) and Buchholz Pancreas (Pancreatic Intraepithelial Neoplasia). The low z-values associated with these datasets indicate a lack of significant correlation between upregulated genes in these datasets, when compared with other datasets. Hence, these two datasets were omitted from further analyses. While we are uncertain about the source of this incongruity, we note that the Buchholz datasets were the only datasets obtained without the use of standard (commercially available) platforms. Another dataset, Logsdon Pancreas, was also omitted due to the low number of genes in the dataset (5,338, compared to an average of 16,652 genes for the rest of the data (Table 1)).

Below, we organize our results and discussion into four discrete sections that include identification of upregulated genes, functional analysis of upregulated genes, identification of a genetic signature for hypomethylation in PC, and identification of potential tissue, serum and matrix metalloproteinase biomarkers in PC.

Identification of upregulated genes

RankProd [24] yields a list of genes ranked by percentage of false positive prediction (pfp) value (see methods). Of the 5590 genes that were upregulated by at least two fold, 827 genes are found to be significantly upregulated when using a pfp threshold of <0.15 [25] (Table S1).

Table 3 provides a list of the top twenty-five ranked genes using the RankProd program. As expected, most genes have well-established associations with pancreatic and other cancers. Some well-known examples include MUC4 [29], CEACAM5/6 [30], S100P [31], CLDN18 [32], KRT19 (CK19) [33] and COLA1/2 [34]. There are, however, some notable exceptions such as AHNK2, CTHRC1, IGHG3 and EPPK1, which do not have a known role in cancer. Hence, these genes can be potential new leads for cancer genes, and are discussed next.

AHNK2 is a significantly upregulated gene in PC (175-fold), but has not been directly associated with any cancer, to our knowledge. The mRNA is reported [35] to be alternatively spliced to produce three isoforms, and the canonical sequence is inferred to be targeted to the nucleus. The AHNK family of scaffold PDZ

Table 1. Pancreatic Cancer Microarray Datasets Included in the Study.

Dataset Name*	Cancer Type	Genes**	Platform	Dataset Summary*
Badea Pancreas	Pancreatic Ductal Adenocarcinoma	19,574	Human Genome U133 Plus 2.0 Array	Paired pancreatic ductal adenocarcinoma (n = 39) and normal pancreas (n = 39) samples from 36 patients were analyzed; three patients were analyzed in duplicate.
Buchholz Pancreas	Pancreatic Ductal Adenocarcinoma	15,725	Human Genome Oligo-Set-Version 2.0 (Operon, Germany)	Eight (8) pancreatic ductal adenocarcinoma and 6 normal pancreatic duct samples were analyzed.
Buchholz Pancreas	Pancreatic Intraepithelial Neoplasia	15,736	Human Genome Oligo-Set-Version 2.0 (Operon, Germany)	Twenty-four (24) pancreatic intraepithelial neoplasia and 6 normal pancreatic duct samples were analyzed.
Grutzmann Pancreas	Pancreatic Ductal Adenocarcinoma	17,782	Human Genome U133A Array, Human Genome U133B Array	Fourteen (14) microdissected pancreatic ductal adenocarcinoma and 11 normal pancreatic duct samples were analyzed. Sample data includes type, age, grade, TNM stage, and sex.
Ishikawa Pancreas	Pancreatic Ductal Adenocarcinoma	17,782	Human Genome U133A Array, Human Genome U133B Array	Twenty-four (24) pancreatic ductal adenocarcinoma and 25 normal pancreatic duct samples were analyzed. Sample data includes type, age, atypical cell proportion, clinical stage, cytological grade, and sex.
Iacobuzio-Donahue Pancreas 2	Pancreatic Ductal Adenocarcinoma	14,361	Non standard	Fourteen (14) pancreatic carcinoma cell lines, 17 primary pancreatic ductal adenocarcinoma samples of various histologies, and 5 normal pancreas samples were analyzed.
Logsdon Pancreas	Pancreatic Adenocarcinoma	5,338	HumanGeneFL Array	Ten microdissected adenocarcinoma, 7 pancreatic cancer cell lines, 5 pancreatitis, and 5 normal pancreas samples were analyzed. Sample data includes type and cell line name.
Pei Pancreas	Pancreatic Ductal Adenocarcinoma	19,574	Human Genome U133 Plus 2.0 Array	Thirty-six (36) pancreatic carcinoma and 16 paired normal samples, for a total of 52 samples, were analyzed. Sample data includes age and sex.
Segara Pancreas	Pancreatic carcinoma	12,684	Human Genome U133A Array	Eleven (11) pancreatic adenocarcinoma samples and six (6) adjacent normal pancreas samples from 12 patients were analyzed.

*As identified by the Oncomine database.

**Number of genes probed.

doi:10.1371/journal.pone.0093046.t001

proteins consists of two large proteins (600–700 kD), AHNAK (desmoyokin) and AHNAK2 [36]. AHNAK has been associated with several muscular diseases, including cardiomyopathy and limb-girdle muscular dystrophy, and this effect is believed to be mediated through its association with the β -subunit of cardiac Ca(v) calcium channel [37]. AHNAK & AHNAK2 have also been shown to be components of the costameric network, associated with linking of the extracellular matrix to the cytoplasmic microfilament system [38]. Experiments on metastatic human tumor cell lines [39] have shown that knockdown of AHNAK resulted in pseudopod retraction, inhibition of cell migration and reversion of mesenchymal-epithelial transition (MET). It is likely that AHNAK and AHNAK2 were both affected by these knockdown experiments. Our results suggest that the family of AHNAK proteins, particularly AHNAK2, merit experimental scrutiny regarding their possible role in carcinogenesis, especially in PC.

CTHRC1 (collagen triple helix containing 1) is a 30 kD secreted protein that has the ability to inhibit collagen matrix

synthesis, and is highly expressed during skin wound healing. Tissue repair and carcinogenesis are linked [40] and CTHRC1 has been associated with a variety of tumors including melanoma [41], breast cancer [42], colorectal cancer [43] and most recently, gastric cancer [44]. However, there has only been one report that links CTHRC1 with PC, where higher expression of CTHRC1 was observed in a screen of solid tumor cell lines including PC [41]. There is evidence that CTHRC1 expression is associated with cancer tissue invasion and metastasis in breast cancer [42] and gastric cancer [44]. Given the high level of upregulation of CTHRC1 (>1,000-fold) that was observed in this study, we hypothesize CTHRC1 to be an excellent candidate for experimental evaluation as a potential biomarker for PC.

IGHG3 (Immunoglobulin heavy constant γ -3) is a secreted antigen binding protein not previously implicated in pancreatic cancer. Our analysis (see next section) indicates that PC is associated with dysfunction of the immune system. IGHG3 is also a component of the top network associated with the cohort of 827 overexpressed genes, which is shown in Figure 1.

Table 2. Pairwise z-values* Indicating Congruency between Upregulated Genes.

Name	Badea ¹	Buch_DAC ²	Buch_Intra ³	Grutzmann ⁴	Iacobuzio ⁵	Ishikawa ⁶	Logsdon ⁷	Pei ⁸
Buch_DAC	0.6							
Buch_Intra	0.2	17.6						
Grutzmann	9.7	-0.2	-0.6					
Iacobuzio	9.7	-0.1	-0.6	5.2				
Ishikawa	4.4	-0.6	-0.6	7.8	2.1			
Logsdon	14.9	0.1	-0.9	8.7	9.5	1.3		
Pei	20.8	-0.6	-1.5	12.4	12.3	4.2	13.9	
Segeera ⁹	17.5	0.7	0.4	7.2	8.2	6.1	11.5	12.7

*A z-value of >1.96 indicates a p-value of <0.05.

¹Badea Pancreas (Pancreatic Ductal Adenocarcinoma).

²Buchholz Pancreas (Pancreatic Ductal Adenocarcinoma).

³Buchholz Pancreas (Pancreatic Intraepithelial Neoplasia).

⁴Grutzmann Pancreas (Pancreatic Ductal Adenocarcinoma).

⁵Iacobuzio-Donahue (Pancreatic Adenocarcinoma).

⁶Ishikawa Pancreas (Pancreatic Ductal Adenocarcinoma).

⁷Logsdon Pancreas (Pancreatic Adenocarcinoma).

⁸Pei Pancreas (Pancreatic Ductal Adenocarcinoma).

⁹Segeera Pancreas (Pancreatic Carcinoma).

doi:10.1371/journal.pone.0093046.t002

Table 3. A List of the 25 Most Highly Ranked Upregulated Genes in Pancreatic Cancer.

Gene*	Gene Function
AHNAK2	Unknown; a component of the costameric network
CDH3	A calcium-dependent cell adhesion molecule.
CEACAM5	Cell surface glycoprotein that plays a role in cell adhesion and in intracellular signaling; binds with another CEACAM to function.
CEACAM6	A cell adhesion molecule; mediates cell adhesion by binding with another CEACAM (-1, -5, and -6 are most common).
CLDN18	Plays a major role in tight junction-specific obliteration of the intercellular space, through calcium-independent cell-adhesion activity
COL11A1	This gene encodes one of the two alpha chains of type XI collagen, a minor fibrillar collagen.
COL1A1	Type 1 collagen is a fibril forming collagen found in most connective tissue; alpha chain one.
COL1A2	Type 1 collagen is a fibril forming collagen found in most connective tissue; alpha chain two.
CTHRC1	May play a role in the cellular response to arterial injury through involvement in vascular remodeling. (secreted)
CTSE	A gastric aspartyl protease that functions as a disulfide-linked homodimer.
EPPK1	Unknown. May play a role in supporting the intermediate filaments
FN1	Fibronectin is involved in cell adhesion and migration processes including embryogenesis, wound healing, blood coagulation, host defense, and metastasis. (secreted)
GPRC5A	Unknown. May be involved with the interaction between retanoic acid and the G protein sigaling pathway.
IGHG3	Unknown; Immunoglobulin heavy chain gamma 3
KRT19	Involved in the organization of myofibers. Together with KRT8, helps to link the contractile apparatus to dystrophin at the costameres of striated muscle.
MMP11	Weakly degrades structural proteins of the ECM.
MUC4	Mucins are glycoprotein that play a role in the protection of epithelial cells. Implicated in renewal and differentiation.
OLFM4	An antiapoptotic factor that promotes tumor growth and is an extracellular matrix glycoprotein that facilitates cell adhesion.
POSTN	Induces cell attachment and spreading and plays a role in cell adhesion.
S100P	S100 proteins are involved in the regulation of a number of cellular processes such as cell cycle progression and differentiation.
SERPINB5	Unknown. Exhibits no serine protease inhibitory activity functions as a tumor suppressor of mammary tumors
SLC6A14	A member of the solute carrier family; transports both neutral and cationic amino acids
VCAN	This protein is involved in cell adhesion, proliferation, migration and angiogenesis and plays a central role in tissue morphogenesis and maintenance
THBS2	A disulfide-linked homotrimeric glycoprotein that mediates cell-to-cell and cell-to-matrix interactions
COL3A1	Pro-alpha1 chain of type III collagen, a fibrillar collagen that is found in extensible connective tissue.

*Gene names are given according to the HUGO Gene Nomenclature Committee (HGNC).

doi:10.1371/journal.pone.0093046.t003

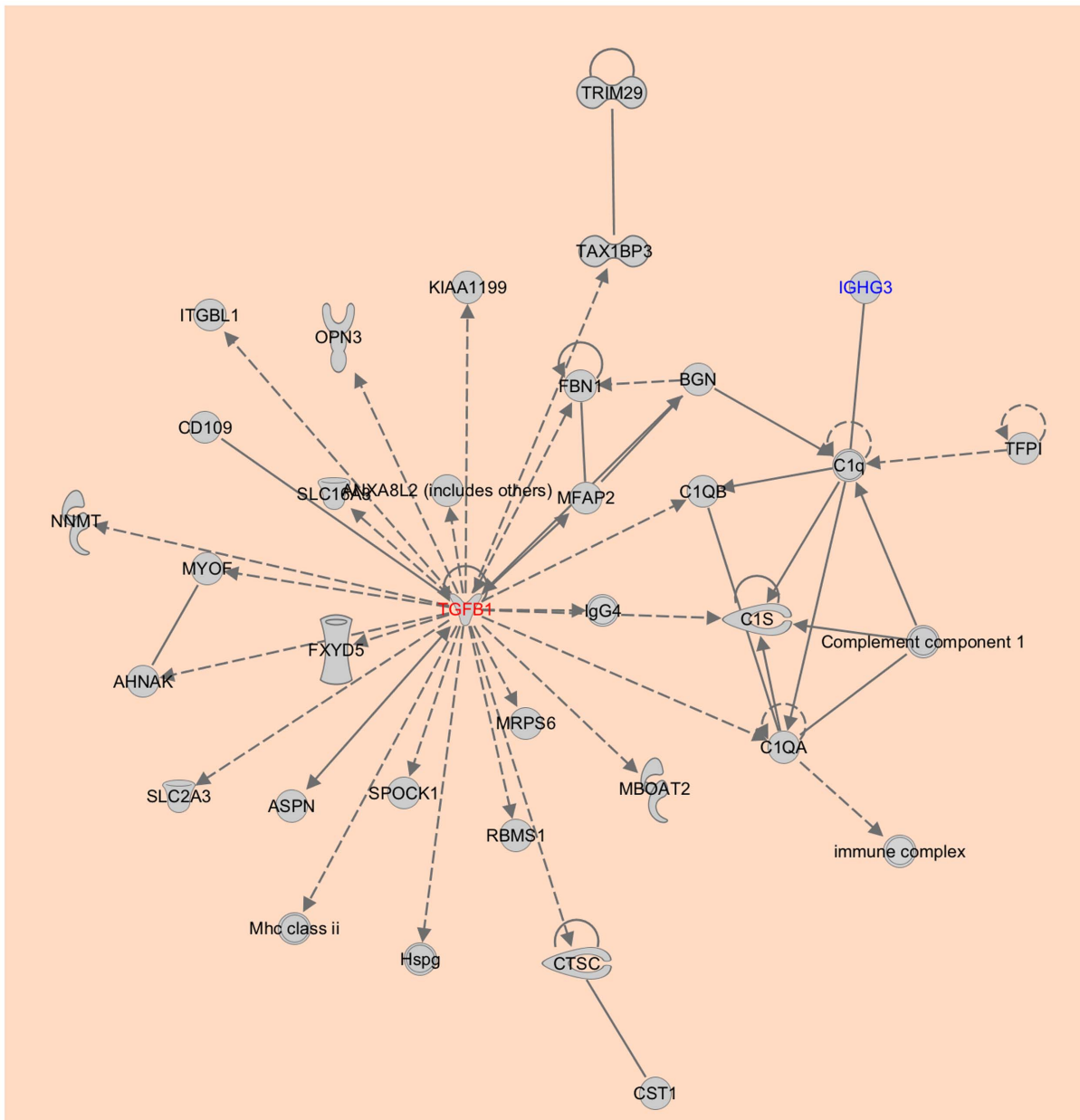


Figure 1. Top scoring network associated with the upregulated genes in PC. TGFB1 forms a hub node in the network. IGHG3 (highlighted in blue color) is one of the top twenty-five genes that is potentially important for pancreatic cancer.
doi:10.1371/journal.pone.0093046.g001

Epiplakin belongs to the plakin family of cytolinker proteins that are associated with the junctional complexes and the cytoskeleton. Epiplakin is rather an unusual plakin in that it consists solely of plakin repeats organized into 13 plakin repeat domains (PRD's) and does not contain a plakin domain characteristic of other plakins. There is evidence to suggest that Epiplakin associates with keratin networks during wound healing [45].

Functional analysis of the upregulated genes

We identified important functions, networks, and pathways relevant to the 827 significantly upregulated genes using IPA (www.ingenuity.com). A comprehensive analysis of the 827 upregulated genes is shown in Table S1.

The most significant biological functions associated with the 827 upregulated genes are cellular movement, cellular growth and proliferation, cell death and survival, cellular development and cell-to-cell signaling and interaction (Figure 2, Table S2). Dysregulation of these functions are associated with cancer and metastasis, reiterating the importance of this geneset to PC. A pathway analysis provided insights into some of the molecular mechanisms important in PC. The five most significant pathways associated with the 827 upregulated genes included integrin signaling (p-value = 1.72×10^{-13}), also observed by Grutzmann et al. [20], granulocyte adhesion and diapedesis (p-value = 4.08×10^{-11}), agranulocyte adhesion and diapedesis (p-value = 9.43×10^{-10}), leukocyte extravasation signaling (p-value = 1.62×10^{-9}), and virus entry via endocytic pathways (p-

value = 1.71×10^{-8}) (Figure 3, Table S3). These results indicated that PC is significantly associated with inflammation and immune mechanisms. In fact, it has been shown that cancer immunosuppression often favors tumor progression and metastasis by constituting an immunosuppressive network in which several tumor-derived soluble factors such as interleukin-10, transforming growth factor beta (TGFB) and vascular endothelial growth factor play central roles [46]. In the top network identified, TGFB1 is the hub gene (Figure 1). TGFB1 encodes a member of the TGFB family of cytokines, which are multifunctional peptides that regulate proliferation, differentiation, adhesion, migration, and other functions in many cell types. This gene has been shown to be frequently upregulated in tumor cells, and is an important target for cancer therapy [47–51].

The second most significant network associated with the upregulated genes is involved in cell cycle, cellular movement, and cancer (Figure 4). In this network, NF- κ B complex acts as a major hub, which functions as a regulator of genes that control cell proliferation and cell survival. Incorrect regulation of NF- κ B has been linked to cancer, inflammatory and autoimmune diseases [52,53]. This network again suggests that PC could be closely correlated with immunological disorder [54,55]. Upregulated NF-

κ B turns on the expression of genes that keep the cell proliferating, and protect the cell from conditions that would otherwise cause it to die via apoptosis. In fact, it has been shown that NF- κ B is constitutively active in various types of human tumors [56–60]. In addition, there are two interesting regulatory modules identified in this network. The first module is made up of two E2F family genes (E2F7, E2F8), ECT2 and RACGAP1. These genes form autoregulatory loops, and regulate each other. Notably, the three genes E2F7, E2F8 and ECT2 constitutively regulate RACGAP1, which binds to Rho GTPases (Figure 4), suggesting that this module functions in the regulation of cytokinesis in a cell cycle-dependent manner. Another module involves the glutathione peroxidase (GPX) family genes that encode an enzyme family with peroxidase activity, whose main biological role is to protect the organism from oxidative damage. Upregulation of GPX family genes may be associated PC and other cancers [61–64], suggesting an important link between oxidatively-induced DNA damage and cancer development.

Identification of the upregulated kinase-encoding genes

Furthermore, we extracted the genes encoding protein kinases from the 827 upregulated genes. Table 4 presents the 26 kinase-

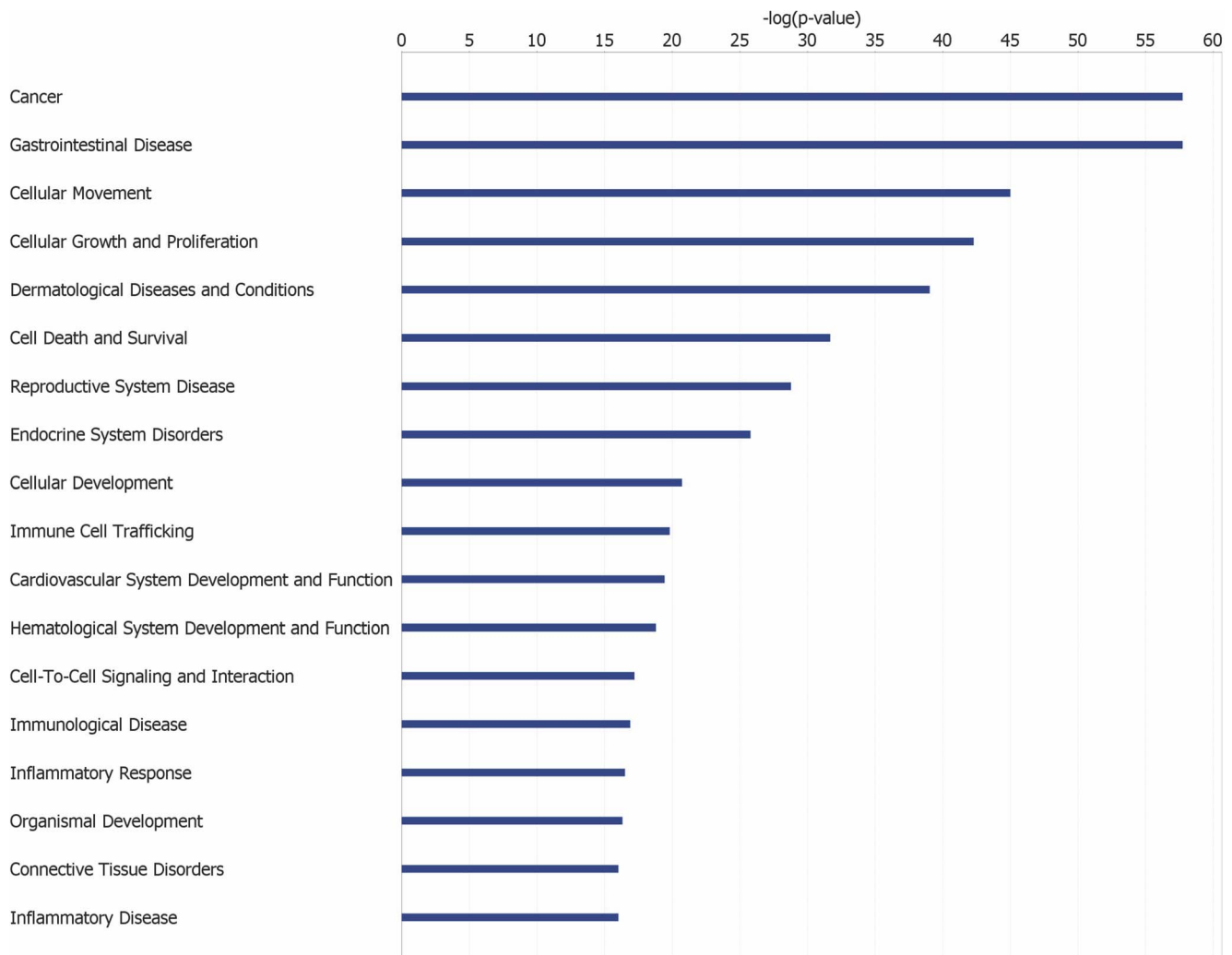


Figure 2. Important biological functions and diseases associated with genes upregulated in PC.

doi:10.1371/journal.pone.0093046.g002

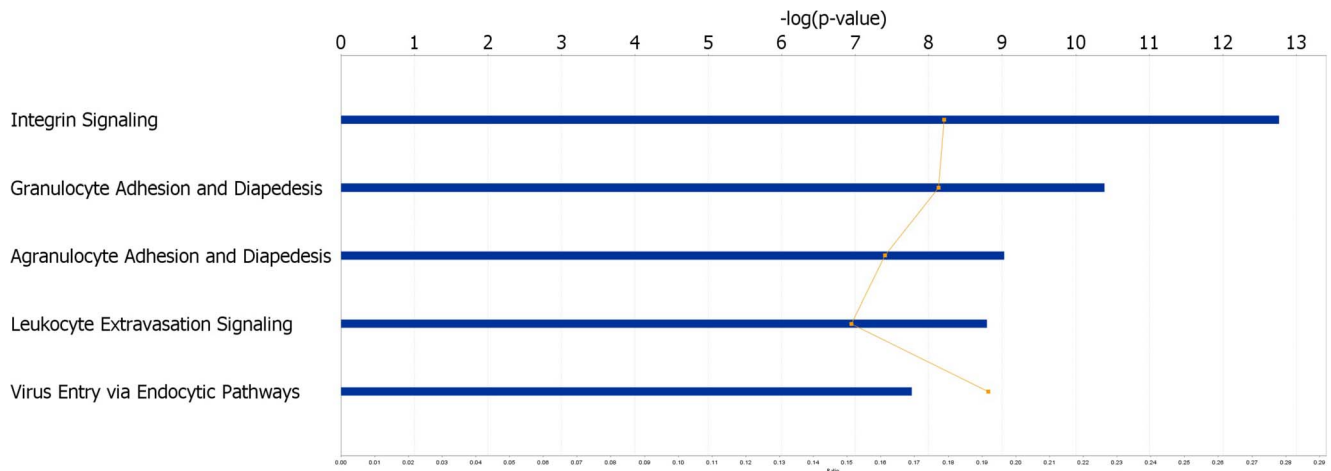


Figure 3. The five most significant pathways associated with genes upregulated in PC are related to inflammation and immune response.

doi:10.1371/journal.pone.0093046.g003

encoding genes. It has been known that many kinase-encoding genes are upregulated in cancer, and development of anticancer drugs that inhibit overexpression of protein kinases has been an active area of research. In fact, effective drugs have already been developed to target some of the protein kinases. For example, CDK1 encodes a member of the Ser/Thr protein kinase family, which is a catalytic subunit of the highly conserved protein kinase complex known as M-phase promoting factor. The protein plays a key role in G1/S and G2/M phase transitions of eukaryotic cell cycle, and the phosphorylation and dephosphorylation of this protein play important regulatory roles in cell cycle control [65]. Some CDK1 kinase inhibitors have been developed for clinical or experimental purposes - AZD 5438, (R)-CR8, (R)-DRF053 dihydrochloride, Kenpaullone, NU 2058, and Ro 3306 (Tocris Bioscience, www.tocris.com), and ZK 304709 and Terameprocol [66].

LCK is a tyrosine-protein kinase that is found inside lymphocytes of the immune system, and involved in immune signaling pathways. Dasatinib, a small-molecule protein tyrosine kinase inhibitor and anticancer drug, can inhibit LCK activity in T-cell activation and proliferation [67,68]. MET is a proto-oncogene that encodes the hepatocyte growth factor receptor protein [69], which possesses tyrosine-protein kinase activity. Abnormal upregulation of MET in cancer often correlates with poor prognosis by triggering tumor growth, angiogenesis that supply the tumor with nutrients, and metastasis. It has been revealed that the MET pathway is one of the most frequently dysregulated pathways in human cancer [70]. A substantial number of MET inhibitors have been studied in clinical trails like AMG-458 (Amgen), PF-04217903 (Pfizer), MK-2461(Merck), ARQ197 (ArQule) etc. [71].

TTK encodes a dual specificity protein kinase with the ability to phosphorylate tyrosine, serine and threonine. TTK kinase is associated with cell proliferation and is essential for the proper attachment of chromosomes to the mitotic spindle. Inhibition of TTK kinase has been shown to correlate with cell death caused by chromosomal missegregations [72]. Several TTK kinase inhibitors have been reported in the literature – Reversine [73], NMS-P715 [74], and MPS1-IN-1 [75].

For some other tyrosine-protein kinases such as LYN, Dasatinib is an effective inhibitor [76]. Of the 26 kinase-encoding genes we identified, some genes have been identified as very promising anticancer targets. For example, BUB1 encoding the mitotic

checkpoint serine/threonine-protein kinase is critical in the establishment of the mitotic spindle checkpoint and chromosome congression. It has been shown that disturbed mitotic checkpoints are a common feature of many human cancers [77]. However, BUB1 expression levels depend on the localization of tumors and their severity [78]. Downregulation of BUB1 resulted in more sarcomas, lymphomas and lung tumors, whereas upregulation of BUB1 caused sarcomas and tumors in the liver [78]. Our result shows that PC is related to the upregulation of BUB1 and we speculate that development of BUB1 inhibitors could provide a new approach to tackling PC.

To sum up, some of the 26 significantly upregulated protein kinase genes in PC could be viable new therapeutic targets for PC. In fact, for the tyrosine-protein kinase genes such as LCK, MET and LYN, which have been found to be frequently overexpressed in human cancer including PC [79], effective tyrosine-protein kinase inhibitors such as Dasatinib, Imatinib, Gefitinib, Erlotinib, and Sunitinib have been developed for anticancer chemotherapy [80].

Towards a genetic signature for hypomethylation in pancreatic cancer

Aberrant hypermethylation of promoter CpG islands is tightly associated with gene silencing, whereas hypomethylation can lead to the upregulation of genes. A recent review [81], discusses genes that have been found to be hypomethylated in PC. With reference to this gene set, we do find a strong correlation between hypomethylation and upregulation; specifically, seven of nine genes mentioned in this review (SERPINB5, CLDN4, SFN, S100P, S100A4, MSLN, and PSCA) are significantly upregulated, with SERPINB5, SFN, S100P, and PSCA being among the 100 most upregulated genes in our analysis (Table S1).

A comprehensive study on aberrant methylation in PC has been performed by Tan et al. [82], who profiled 1505 CpG sites across 807 genes. Initial investigations yielded a list of 63 genes with CpG site hypomethylation and increased mRNA expression. Somewhat unexpectedly, the authors also found a similar number of genes with CpG site hypomethylation and decreased mRNA expression. Upon further experimentation, 35 of the 63 genes were identified by the authors as candidate genes that are regulated by hypomethylation in PC. We find that eight of the 35 candidate genes (ID1, MMP7, MST1R, NBL1, PHLDA2, PLAT, PLAUR

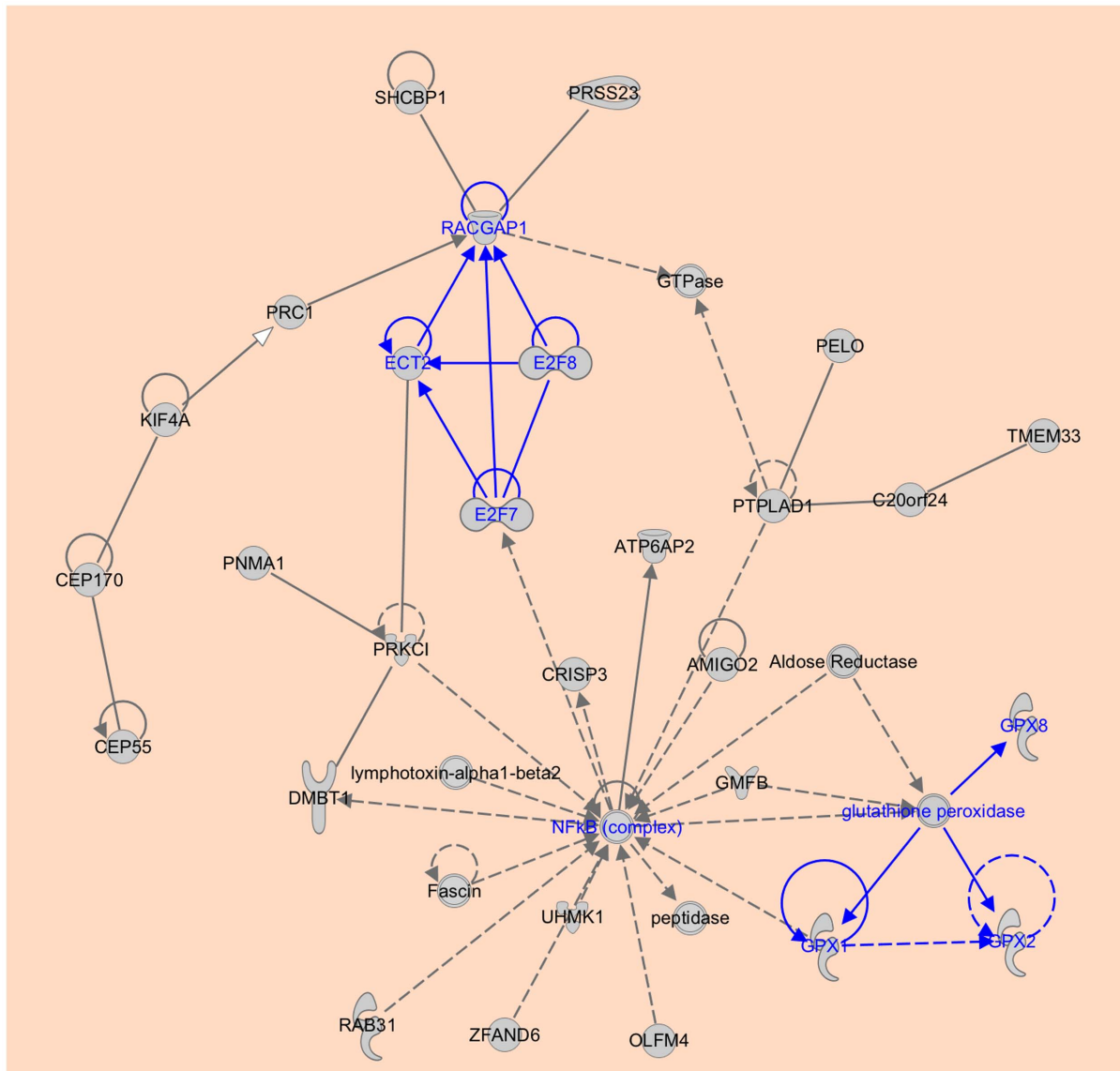


Figure 4. The second most significant network associated with the upregulated genes in PC. A major hub node NF- κ B complex, and two new regulatory modules are highlighted in blue color. One module is made up of two E2F family genes (E2F7, E2F8), ECT2 and RACGAP1; and another module is made up of several GPX family genes.
doi:10.1371/journal.pone.0093046.g004

and SFN), and a further 8 (IL8, SPP1, CLDN4, MMP1, ARHGDI5, NQO1, ITGB4, SERPINB5, and TFF1) from the original list of 63 genes are also significantly upregulated in our study.

To summarize, twenty-two genes (MUC4, SERPINB5, CLDN4, SFN, TFF1, S100P, S100A4, MMP1, MMP7, MSLN, PSCA, ID1, MST1R, NBL1, PHLDA2, PLAT, PLAUR, IL8, SPP1, ARHGDI5, NQO1, and ITGB4) are significantly upregulated in our analysis, and there is experimental evidence [81,82] to suggest that this upregulation is due to hypomethylation. Thus, these genes will contribute towards a growing list of candidates including MUC4 [83] that describe a putative genetic signature for hypomethylation in pancreatic cancer (Table 5). Such a genetic signature could prove to be useful in the early detection of PC, in a manner analogous to the clinical use of aberrant methylation of CCND2 [84] in PC. Since there is an emerging consensus that ‘epigenetic chaos’ promoted changes in gene expression and,

ultimately, leads to cancer [85], it is quite likely that many of the genes found to be significantly upregulated (Table S1) are hypomethylated in PC. Of the 22 genes, IPA analysis reveals that 11 genes have a known association with PC (Table 5).

Potential biomarkers among upregulated genes

Tumor tissue protein biomarkers. An observation often reported in literature is the discrepancy between the level of expression of a protein and that of its transcript for a given type of cell [86]. Nonetheless, we find about 70% of thirty two tumor tissue protein biomarkers identified in two recent reviews [87,88] were found to be upregulated >2 -fold in our analysis. Among those significantly upregulated ($p < 0.15$) were a cluster of genes associated with the actin microfilament, IGALS1 (galectin-1), ACTN4 (actinin-4), PLS1 (plastin-1), TPM2 (tropomyosin β), CFL1 (cofilin-1), ENO1 (α -enolase), and MSN (moesin). Most of these proteins are known actin-binding proteins that can modulate

Table 4. Twenty-six Kinase-encoding Genes Upregulated in Pancreatic Cancer.

Gene*	Description
ACVR1	Activin A receptor, type I
BUB1	Budding uninhibited by benzimidazoles 1 homolog (yeast)
BUB1B	Budding uninhibited by benzimidazoles 1 homolog beta (yeast)
CDK1	Cell division cycle 2, G1 to S and G2 to M
DYRK2	Dual-specificity tyrosine-(Y)-phosphorylation regulated kinase 2
EPHA4	EPH receptor A4
IRAK3	Interleukin-1 receptor-associated kinase 3
LCK	lymphocyte-specific protein tyrosine kinase
LYN	V-yes-1 Yamaguchi sarcoma viral related oncogene homolog
MAP4K4	Mitogen-activated protein kinase 4
MELK	Maternal embryonic leucine zipper kinase
MET	Met proto-oncogene (hepatocyte growth factor receptor)
MST1R	Macrophage stimulating 1 receptor (c-met-related tyrosine kinase)
MST4	Serine/threonine protein kinase MST4
NEK2	NIMA (never in mitosis gene a)-related kinase 2
NUAK1	NUAK family, SNF1-like kinase, 1
PBK	PDZ binding kinase
PRKCI	Protein kinase C, iota
PTK6	PTK6 protein tyrosine kinase 6
STK17B	Serine/threonine kinase 17b
STK24	Serine/threonine kinase 24 (STE20 homolog, yeast)
STYK1	Serine/threonine/tyrosine kinase 1
TNIK	TRAF2 and NCK interacting kinase
TRIB2	Tribbles homolog 2 (Drosophila)
TTK	TTK protein kinase
UHMK1	U2AF homology motif (UHM) kinase 1

*Gene names are given according to the HUGO Gene Nomenclature Committee (HGNC).

doi:10.1371/journal.pone.0093046.t004

the actin microfilament, or modulate its environment with the plasma membrane.

Other suggested tumor tissue protein biomarkers [87,88] significantly upregulated in our analysis include SFN, AGR2, LGALS1, LGALS3, THBS2, & TGFBI, and four members of the S100 family, S100A6, S100A10, S100A11, and S100A2 [89]. We find three additional members of the S100 family, S100A4, S100A16 and S100P were also significantly upregulated (Table S1). The S100 family of low molecular weight calcium binding proteins have strong associations with cancer [90], and several of them have been used as markers in melanoma and other cancers. It should be noted that S100P is one of the most upregulated genes in our analysis ($> \times 10^6$). It has recently been proposed that S100P be used as a protein biomarker for intraductal papillary mucinous neoplasms (IPMN) of the pancreas [91], and for pancreatic adenocarcinoma [92].

Serum protein biomarkers. Early diagnosis of pancreatic cancer is essential in order to improve the poor prognosis associated with PC. Serum biomarkers offer a very attractive and non-invasive solution, and are thus highly sought after [14]. However, there is a paucity of serum biomarkers for PC [15], with the carbohydrate biomarker CA 19-9 being the most widely used.

Since serum protein biomarkers such as CA-125 may be cleaved and released in PC [93] a correlation between serum biomarkers and mRNA expression is not necessarily expected (though in the case of CA-125, there is evidence that it is overexpressed as well [93]). Nevertheless, we sought to investigate whether any of the proposed serum protein biomarkers in the recent literature [3] were upregulated in pancreatic cancer at the level of mRNA. Somewhat to our surprise, we found that about one-third of the corresponding genes, C3, B2M, C1QB, CD9, TIMP1, PGK1, SERPINA1, APOE, AGR2, APOC1 & SPP1, were significantly upregulated in our analysis. These results indicate our cohort of 827 significantly upregulated genes also represent an enriched pool of candidate serum protein biomarkers. The commercial availability of many human antibodies raises the intriguing possibility of performing a systematic screen of serum, to detect for protein products of significantly upregulated genes in our analysis. While individual biomarkers may suffer from issues of sensitivity and specificity [14], the promise is that with a large number of biomarkers, distinctive signatures are likely to emerge, that correlate with diagnosis and prognosis.

Matrix metalloproteinase biomarkers. Matrix metalloproteases represent the most prominent family of proteinases associated with tumorigenesis [94]. In our analysis, we found that seven matrix metalloproteases (MMPs) and six proteases from a related family “a disintegrin and metalloprotease” (ADAMs) to be significantly upregulated (Table 6). Three of these (MMP9, ADAM9 and ADAM10) were also found to be upregulated by Grutzman *et al.* [20].

Matrix metalloproteases are a family of zinc-dependent proteases that have the capacity to degrade virtually every component of the extracellular matrix (ECM). Tumor cells overexpress these proteases in order to degrade the basement membrane and invade the surrounding tissue. This activity is also required for the intravasation and extravasation events in metastasis. MMP substrates also include non-ECM molecules, ranging from growth factor precursors and cell surface adhesion molecules to angiogenic inhibitor precursors [95]. MMPs have also been implicated in the epithelial to mesenchymal transition (EMT) [96]. While MMPs have well-recognized roles in the late stage of tumor progression, invasion, and metastasis, emerging evidence suggests that the role of MMPs in tumorigenesis is more complex [97].

One of the more promising and exciting applications of MMPs in human cancers is as potential cancer biomarkers, both diagnostic and prognostic. MMP-2, MMP-7 and MMP-9 are among the most well studied matrix metalloproteases in PC [98]. MMP-9 expression has been linked to worse prognosis, and it also significantly correlated with tumor expression and distant metastasis [99]. Active MMP-2 levels are upregulated in the pancreatic juice of patients with cancer (100%) as compared with patients with chronic pancreatitis (2%) or normal controls (0%) [100]. Similarly, plasma as well as tumor tissues from patients with pancreatic ductal adenocarcinoma have significantly elevated MMP-7 levels, which may predict shortened survival of patients [101].

As expected, MMP-2, MMP-7 and MMP-9 are all significantly upregulated in our study. However, another matrix metalloprotease, MMP-11, is the most highly upregulated MMP, with an average $>10,000$ fold overexpression in PC. MMP-11 induction in adipose tissue has been linked to cancer progression [102] and MMP-11 has been associated with tumor progression in pulmonary cancer [103], head and neck carcinoma [104] and breast carcinoma [105]. MMP-11 is known to cleave IGF binding proteins, which regulate the bioavailability of insulin-like growth

Table 5. A Putative Genetic Signature of Hypomethylated Genes in Pancreatic Cancer.

Gene	pfp value	Log (2) value	Reference*
MUC4 [§]	0.000	6.28	Zhu et al., 2011 [83]
SERPIN5	0.000	8.34	Sato et al., 2003 [110]; Fitzgerald et al., 2003 [111]; Ohike et al., 2003 [112]
CLDN4	0.103	2.67	Sato et al., 2003 [110], Tan et al., 2009 [82] (Gastric sarcoma Kwon et al., 2011 [113])
SFN	0.000	7.91	Sato et al., 2003 [110], Tan et al., 2009 [82] (Lung adenocarcinoma Shiba-Ishii et al., 2012 [114])
TFF1 [§]	0.001	7.27	Tan et al., 2009 [82] (Prostrate cancer Vestergaard et al., 2010 [115])
S100P [§]	0.000	24.25	Sato et al., 2003 [110]
S100A4	0.004	3.12	Rosty et al., 2002 [116]
MMP1	0.003	4.63	Tan et al., 2009 [82]
MMP7 [§]	0.011	4.07	Tan et al., 2009 [82]
MSLN [§]	0.026	4.80	Sato et al., 2003 [110] (Mesothelioma Nelson et al., 2011 [117])
PSCA [§]	0.001	7.78	Sato et al., 2003 [110]
ID1 [§]	0.151	2.06	Tan et al., 2009 [82]
MST1R	0.025	3.40	Tan et al., 2009 [82]
NBL1	0.035	2.82	Tan et al., 2009 [82]
PHLDA2	0.000	5.22	Tan et al., 2009 [82] (Osteosarcoma Li et al., 2008 [118])
PLAT [§]	0.014	3.34	Tan et al., 2009 [82]
PLAUR [§]	0.014	3.18	Tan et al., 2009 [82]
IL8	0.007	4.21	Tan et al., 2009 [82] (Colorectal adenocarcinoma Dimberg et al., 2012 [119])
SPP1 [§]	0.044	2.14	Tan et al., 2009 [82] (Liver fibrosis Komatsu et al., 2012 [120])
ARHGDI5	0.021	2.40	Tan et al., 2009 [82]
NQO1	0.000	6.19	Tan et al., 2009 [82]
ITGB4 [§]	0.014	3.06	Tan et al., 2009 [82]

*References for hypomethylation in other cancers are given in parenthesis.

[§]IPA analysis indicates a known association with PC.

doi:10.1371/journal.pone.0093046.t005

Table 6. Matrix Metalloproteinases Upregulated in Pancreatic Cancer.

Gene	pfp value	Log (2) value
MMP11	0.00	15.71
MMP12	0.00	4.49
MMP1	0.00	4.63
MMP7	0.01	4.07
MMP2	0.03	2.79
MMP28	0.07	2.37
MMP9	0.15	1.87
ADAM8	0.01	2.91
ADAM9	0.05	2.56
ADAM12	0.06	2.34
ADAMTS6	0.07	2.97
ADAMTS12	0.08	2.54
ADAM28	0.10	2.50
ADAM10	0.16	1.67

doi:10.1371/journal.pone.0093046.t006

factors (IGFs). We also found two other MMPs known to cleave IGF binding proteins, MMP-1 and MMP-2, as well as ADAM12, to be significantly upregulated. While the failure of MMP inhibitors in clinical trials has been disappointing [106], our results indicate that MMPs continue to be attractive therapeutic targets for PC.

Tumor tissue heterogeneity. Most cancers are believed to originate through of a process of Darwinian evolution occurring among the cells within the microenvironments provided by the tissues of a multicellular organism. It has become increasingly clear that this process can give rise to tumor tissue heterogeneity [107], with distinct populations of cancer cells predominating in pancreatic and other tumors [16] [108]. For example, this also provides a mechanism for the development of drug resistance, whereby a minor drug resistant subclone in the original tumor becomes dominant after treatment [109]. In this context it is possible that for the datasets in our study, the number of cancer genomes sampled was higher than the number of patient samples. The task of identifying and validating diagnostic and prognostic biomarkers is likely to be complicated by the existence of tumor heterogeneity.

Sample heterogeneity. Most microarray datasets specifically cited the patient samples as being from either pancreatic adenocarcinoma (PAC) or pancreatic ductal adenocarcinoma (PDAC) (Table 1). These samples could still contain contaminants from the desmoplasia, particularly in studies where microdissection was not used (Table 1), contributing to sample heterogeneity.

If significant contamination of tumor samples from immune components of the desmoplasia has occurred, it can have an impact on the association we found by IPA analysis between PC and inflammation/immune mechanisms. It should be noted that there is also support for an association between PC and inflammation from the literature (which we have cited previously). A second source of sample heterogeneity is the type of PC. One study (Pei Pancreas) did not specifically mention PAC or PDAC under the dataset summary, and thus could conceivably contain samples from other types of PC, although PDAC accounts for over 90% of the cases of PC.

The broad concordance observed between the microarray datasets (see Congruency between microarray datasets) suggests that issues related to sample heterogeneity (as well as other sources of variation between the microarray datasets) were not a major complicating factor in this meta-analysis. This observation also strengthens the case for investigating differentially regulated genes as putative biomarkers for PC.

Conclusions

Meta-analysis of multiple microarray datasets can yield more reliable and comprehensive results than using a single dataset, because the former has increased statistical power and generalizability. In the present study, we performed a meta-analysis of nine PC datasets and identified 827 genes that are significantly upregulated in pancreatic cancer. The two most important biological networks associated with these genes have TGF β 1 and NF- κ B as major hubs. A pathway analysis indicates that PC is significantly associated with inflammation and immune mechanism.

References

- Jemal A, Siegel R, Ward E, Hao Y, Xu J, et al. (2009) Cancer statistics, 2009. *CA Cancer J Clin* 59: 225–249.
- Howlander N, Noone AM, Krapcho M, Neyman N, Aminou R, et al. (2011) SEER Cancer Statistics Review, 1975–2009.
- Chakraborty S, Baine MJ, Sasson AR, Batra SK (2011) Current status of molecular markers for early detection of sporadic pancreatic cancer. *Biochim Biophys Acta* 1815: 44–64.
- Stratton MR (2011) Exploring the genomes of cancer cells: progress and promise. *Science* 331: 1553–1558.
- Jones S, Zhang X, Parsons DW, Lin JC, Leary RJ, et al. (2008) Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science* 321: 1801–1806.
- Greenman C, Stephens P, Smith R, Dalgleish GL, Hunter C, et al. (2007) Patterns of somatic mutation in human cancer genomes. *Nature* 446: 153–158.
- Schinzl AC, Hahn WC (2008) Oncogenic transformation and experimental models of human cancer. *Front Biosci* 13: 71–84.
- Beerenwinkel N, Antal T, Dingli D, Traulsen A, Kinzler KW, et al. (2007) Genetic progression and the waiting time to cancer. *PLoS Comput Biol* 3: e225.
- Futreal PA, Coin L, Marshall M, Down T, Hubbard T, et al. (2004) A census of human cancer genes. *Nat Rev Cancer* 4: 177–183.
- Touw IP, Erkeland SJ (2007) Retroviral insertion mutagenesis in mice as a comparative oncogenomics tool to identify disease genes in human leukemia. *Mol Ther* 15: 13–19.
- Stuart D, Sellers WR (2009) Linking somatic genetic alterations in cancer to therapeutics. *Curr Opin Cell Biol* 21: 304–310.
- Ko JM, Fisher DE (2011) A new era: melanoma genetics and therapeutics. *J Pathol* 223: 241–250.
- Marson CM (2009) Histone deacetylase inhibitors: design, structure-activity relationships and therapeutic implications for cancer. *Anticancer Agents Med Chem* 9: 661–692.
- Bunger S, Laubert T, Roblick UJ, Habermann JK (2011) Serum biomarkers for improved diagnostic of pancreatic cancer: a current overview. *J Cancer Res Clin Oncol* 137: 375–389.
- Xu Qao Advances in early diagnosis and therapy of pancreatic cancer.
- Yachida S, Jones S, Bozic I, Antal T, Leary R, et al. (2010) Distant metastasis occurs late during the genetic evolution of pancreatic cancer. *Nature* 467: 1114–1117.
- Goonetilleke KS, Siriwardena AK (2007) Systematic review of carbohydrate antigen (CA 19-9) as a biochemical marker in the diagnosis of pancreatic cancer. *Eur J Surg Oncol* 33: 266–270.
- Ramasamy A, Mondry A, Holmes CC, Altman DG (2008) Key issues in conducting a meta-analysis of gene expression microarray datasets. *PLoS Med* 5: e184.
- Irizarry RA, Warren D, Spencer F, Kim IF, Biswal S, et al. (2005) Multiple-laboratory comparison of microarray platforms. *Nat Methods* 2: 345–350.
- Grutzmann R, Boriss H, Ammerpohl O, Luttgens J, Kalthoff H, et al. (2005) Meta-analysis of microarray data on pancreatic cancer defines a set of commonly dysregulated genes. *Oncogene* 24: 5079–5088.
- Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, et al. (2004) ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia* 6: 1–6.
- Rhodes DR, Kalyana-Sundaram S, Mahavisno V, Varambally R, Yu J, et al. (2007) OncoPrint 3.0: Genes, Pathways, and Networks in a Collection of 18,000 Cancer Gene Expression Profiles. *NEO* 9: 166–180.
- Smid M, Dorsers LCJ, Jenster G (2003) Venn Mapping: clustering of heterologous microarray data based on the number of co-occurring differentially expressed genes. *Bioinformatics* 19: 2065–2071.
- Hong F, Breitling R, McEntee CW, Wittner BS, Nemhauser JL, et al. (2006) RankProd: a bioconductor package for detecting differentially expressed genes in meta-analysis. *Bioinformatics* 22: 2825–2827.
- Laing E, Smith CP (2010) RankProdIt: A web-interactive Rank Products analysis tool. *BMC Res Notes* 3: 221.
- Breitling R, Armengaud P, Amtmann A, Herzyk P (2004) Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. *FEBS Lett* 573: 83–92.
- Dudley JT, Tibshirani R, Deshpande T, Butte AJ (2009) Disease signatures are robust across tissues and experiments. *Mol Syst Biol* 5: 307.
- Hong F, Breitling R (2008) A comparison of meta-analysis methods for detecting differentially expressed genes in microarray experiments. *Bioinformatics* 24: 374–382.
- Yonezawa S, Higashi M, Yamada N, Yokoyama S, Kitamoto S, et al. (2011) Mucins in human neoplasms: clinical pathology, gene expression and diagnostic application. *Pathol Int* 61: 697–716.
- Blumenthal RD, Leon E, Hansen HJ, Goldenberg DM (2007) Expression patterns of CEACAM5 and CEACAM6 in primary and metastatic cancers. *BMC Cancer* 7: 2.

Among the list of candidate cancer genes uncovered by this study are four highly expressed genes not previously associated with PC, and twenty-six kinase genes. Kinases have been attractive targets in combating cancer, and in fact, effective therapeutics have already been developed for several kinases in our list. Importantly, this study also revealed potential biomarkers for pancreatic cancer. Such biomarkers are in urgent need, given the poor prognosis after (the normally late) diagnosis of PC. Towards this end, we have also developed a putative genetic signature for hypomethylated genes in PC. The identification of candidate cancer genes and putative biomarkers for pancreatic cancer provide new opportunities for early diagnosis and treatment of PC.

Supporting Information

Table S1 The 827 genes significantly upregulated in pancreatic cancer.

(XLSB)

Table S2 The most significant biological functions and diseases associated with the 827 upregulated genes.

(XLS)

Table S3 The five most significant pathways associated with the 827 upregulated genes.

(XLS)

Author Contributions

Conceived and designed the experiments: NCWG CG XW. Performed the experiments: NCWG XW LL. Analyzed the data: NCWG CG XW LL. Contributed reagents/materials/analysis tools: NCWG XW. Wrote the paper: NCWG XW.

31. Jiang H, Hu H, Tong X, Jiang Q, Zhu H, et al. (2012) Calcium-binding protein S100P and cancer: mechanisms and clinical relevance. *J Cancer Res Clin Oncol* 138: 1–9.
32. Tanaka M, Shibahara J, Fukushima N, Shinozaki A, Umeda M, et al. (2011) Claudin-18 is an early-stage marker of pancreatic carcinogenesis. *J Histochem Cytochem* 59: 942–952.
33. Jain R, Fischer S, Serra S, Chetty R (2010) The use of cytokeratin 19 (CK19) immunohistochemistry in lesions of the pancreas, gastrointestinal tract, and liver. *Appl Immunohistochem Mol Morphol* 18: 9–15.
34. Provenzano PP, Eliceiri KW, Campbell JM, Inman DR, White JG, et al. (2006) Collagen reorganization at the tumor-stromal interface facilitates local invasion. *BMC Med* 4: 38.
35. Magrane M, Consortium U (2011) UniProt Knowledgebase: a hub of integrated protein data. Database (Oxford) 2011: bar009.
36. Komuro A, Masuda Y, Kobayashi K, Babbitt R, Gunel M, et al. (2004) The AHNAKs are a class of giant propeller-like proteins that associate with calcium channel proteins of cardiomyocytes and other cells. *Proc Natl Acad Sci U S A* 101: 4053–4058.
37. Haase H, Alvarez J, Petzhold D, Doller A, Behlke J, et al. (2005) Ahnak is critical for cardiac Ca(V)1.2 calcium channel function and its beta-adrenergic regulation. *FASEB J* 19: 1969–1977.
38. Marg A, Haase H, Neumann T, Kouno M, Morano I (2010) AHNAK1 and AHNAK2 are costameric proteins: AHNAK1 affects transverse skeletal muscle fiber stiffness. *Biochem Biophys Res Commun* 401: 143–148.
39. Shankar J, Messenberg A, Chan J, Underhill TM, Foster LJ, et al. (2010) Pseudopodial actin dynamics control epithelial-mesenchymal transition in metastatic cancer cells. *Cancer Res* 70: 3780–3790.
40. Beachy PA, Karhadkar SS, Berman DM (2004) Tissue repair and stem cell renewal in carcinogenesis. *Nature* 432: 324–331.
41. Tang L, Dai DL, Su M, Martinka M, Li G, et al. (2006) Aberrant expression of collagen triple helix repeat containing 1 in human solid cancers. *Clin Cancer Res* 12: 3716–3722.
42. Turashvili G, Bouchal J, Baumforth K, Wei W, Dziechciarkova M, et al. (2007) Novel markers for differentiation of lobular and ductal invasive breast carcinomas by laser microdissection and microarray analysis. *BMC Cancer* 7: 55.
43. Chiu ST, Hsieh FJ, Chen SW, Chen CL, Shu HF, et al. (2005) Clinicopathologic correlation of up-regulated genes identified using cDNA microarray and real-time reverse transcription-PCR in human colorectal cancer. *Cancer Epidemiol Biomarkers Prev* 14: 437–443.
44. Wang P, Wang YC, Chen XY, Shen ZY, Cao H, et al. (2012) CTHRC1 is upregulated by promoter demethylation and transforming growth factor-beta1 and may be associated with metastasis in human gastric cancer. *Cancer Sci* 153: 1005–1011.
45. Ishikawa K, Sumiyoshi H, Matsuo N, Takeo N, Goto M, et al. (2010) Epiplakin accelerates the lateral organization of keratin filaments during wound healing. *J Dermatol Sci* 60: 95–104.
46. Kim R, Emi M, Tanabe K (2006) Cancer immunosuppression and autoimmune disease: beyond immunosuppressive networks for tumour immunity. *Immunology* 119: 254–264.
47. Leivonen SK, Kahari VM (2007) Transforming growth factor-beta signaling in cancer invasion and metastasis. *Int J Cancer* 121: 2119–2124.
48. Ganapathy V, Ge R, Grazioli A, Xie W, Banach-Petrosky W, et al. (2010) Targeting the Transforming Growth Factor-beta pathway inhibits human basal-like breast cancer metastasis. *Mol Cancer* 9: 122.
49. Ray DM, Myers PH, Painter JT, Hoenerhoff MJ, Olden K, et al. (2012) Inhibition of transforming growth factor-beta-activated kinase-1 blocks cancer cell adhesion, invasion, and metastasis. *Br J Cancer* 107: 129–136.
50. Melisi D, Ishiyama S, Scwab GM, Fleming JB, Xia Q, et al. (2008) LY2109761, a novel transforming growth factor beta receptor type I and type II dual inhibitor, as a therapeutic approach to suppressing pancreatic cancer metastasis. *Mol Cancer Ther* 7: 829–840.
51. Korpala M, Yan J, Lu X, Xu S, Lerit DA, et al. (2009) Imaging transforming growth factor-beta signaling dynamics and therapeutic response in breast cancer bone metastasis. *Nat Med* 15: 960–966.
52. May MJ, Ghosh S (1998) Signal transduction through NF-kappa B. *Immunol Today* 19: 80–88.
53. Karin M, Staudt L (2010) NF-kappa B: a network hub controlling immunity, inflammation, and cancer. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press. 271 p.
54. Elinav E, Nowarski R, Thaiss CA, Hu B, Jin C, et al. (2013) Inflammation-induced cancer: crosstalk between tumours, immune cells and microorganisms. *Nat Rev Cancer* 13: 759–771.
55. Lesina M, Kurkowski MU, Ludes K, Rose-John S, Treiber M, et al. (2011) Stat3/Socs3 activation by IL-6 transsignaling promotes progression of pancreatic intraepithelial neoplasia and development of pancreatic cancer. *Cancer Cell* 19: 456–469.
56. Yoon J, Cho SJ, Ko YS, Park J, Shin DH, et al. (2013) A synergistic interaction between transcription factors nuclear factor-kappaB and signal transducers and activators of transcription 3 promotes gastric cancer cell migration and invasion. *BMC Gastroenterol* 13: 29.
57. Van Laere SJ, Van der Auwera I, Van den Eynden GG, Elst HJ, Weyler J, et al. (2006) Nuclear factor-kappaB signature of inflammatory breast cancer by cDNA microarray validated by quantitative real-time reverse transcription-PCR, immunohistochemistry, and nuclear factor-kappaB DNA-binding. *Clin Cancer Res* 12: 3249–3256.
58. Song JK, Jo MR, Park MH, Song HS, An BJ, et al. (2012) Cell growth inhibition and induction of apoptosis by snake venom toxin in ovarian cancer cell via inactivation of nuclear factor kappaB and signal transducer and activator of transcription 3. *Arch Pharm Res* 35: 867–876.
59. Peant B, Diallo JS, Lessard L, Delvoe N, Le Page C, et al. (2007) Regulation of IkappaB kinase epsilon expression by the androgen receptor and the nuclear factor-kappaB transcription factor in prostate cancer. *Mol Cancer Res* 5: 87–94.
60. Lessard L, Begin LR, Gleave ME, Mes-Masson AM, Saad F (2005) Nuclear localisation of nuclear factor-kappaB transcription factors in prostate cancer: an immunohistochemical study. *Br J Cancer* 93: 1019–1023.
61. Yuzhalin AE, Kutikhin AG (2012) Inherited variations in the SOD and GPX gene families and cancer risk. *Free Radic Res* 46: 581–599.
62. Erdem O, Eken A, Akay C, Arsova-Sarafinovska Z, Matevska N, et al. (2012) Association of GPX1 polymorphism, GPX activity and prostate cancer risk. *Hum Exp Toxicol* 31: 24–31.
63. Cox DG, Tamimi RM, Hunter DJ (2006) Gene x Gene interaction between MNSD and GPX-1 and breast cancer risk: a nested case-control study. *BMC Cancer* 6: 217.
64. Liu J, Du J, Zhang Y, Sun W, Smith BJ, et al. (2006) Suppression of the malignant phenotype in pancreatic cancer by overexpression of phospholipid hydroperoxide glutathione peroxidase. *Hum Gene Ther* 17: 105–116.
65. Enserink JM, Kolodner RD (2010) An overview of Cdk1-controlled targets and processes. *Cell Div* 5: 11.
66. Zhu F, Shi Z, Qin C, Tao L, Liu X, et al. (2012) Therapeutic target database update 2012: a resource for facilitating target-oriented drug discovery. *Nucleic Acids Res* 40: D1128–36.
67. Schade AE, Schieven GL, Townsend R, Jankowska AM, Susulic V, et al. (2008) Dasatinib, a small-molecule protein tyrosine kinase inhibitor, inhibits T-cell activation and proliferation. *Blood* 111: 1366–1377.
68. Lee KC, Ouwehand I, Giannini AL, Thomas NS, Dibb NJ, et al. (2010) Lck is a key target of imatinib and dasatinib in T-cell activation. *Leukemia* 24: 896–900.
69. Bottaro DP, Rubin JS, Falletto DL, Chan AM, Kmieciak TE, et al. (1991) Identification of the hepatocyte growth factor receptor as the c-met proto-oncogene product. *Science* 251: 802–804.
70. Liu X, Newton RC, Scherle PA (2010) Developing c-MET pathway inhibitors for cancer therapy: progress and challenges. *Trends Mol Med* 16: 37–45.
71. Underiner TL, Herberich T, Miknyoczki SJ (2010) Discovery of small molecule c-Met inhibitors: Evolution and profiles of clinical candidates. *Anticancer Agents Med Chem* 10: 7–27.
72. Vijay Kumar D, Hoarau C, Bursavich M, Slattum P, Gerrish D, et al. (2012) Lead optimization of purine based orally bioavailable Mps1 (TTK) inhibitors. *Bioorg Med Chem Lett* 22: 4377–4385.
73. Santaguida S, Tighe A, D'Alise AM, Taylor SS, Musacchio A (2010) Dissecting the role of MPS1 in chromosome biorientation and the spindle checkpoint through the small molecule inhibitor reversine. *J Cell Biol* 190: 73–87.
74. Caldarelli M, Angiolini M, Disingrini T, Donati D, Guanci M, et al. (2011) Synthesis and SAR of new pyrazolo[4,3-h]quinazoline-3-carboxamide derivatives as potent and selective MPS1 kinase inhibitors. *Bioorg Med Chem Lett* 21: 4507–4511.
75. Kwiatkowski N, Jelluma N, Filippakopoulos P, Soundararajan M, Manak MS, et al. (2010) Small-molecule kinase inhibitors provide insight into Mps1 cell cycle function. *Nat Chem Biol* 6: 359–368.
76. Choi YL, Bocanegra M, Kwon MJ, Shin YK, Nam SJ, et al. (2010) LYN is a mediator of epithelial-mesenchymal transition and a target of dasatinib in breast cancer. *Cancer Res* 70: 2296–2306.
77. Williams BR, Amon A (2009) Aneuploidy: cancer's fatal flaw? *Cancer Res* 69: 5289–5291.
78. King RW (2008) When 2+2=5: the origins and fates of aneuploid and tetraploid cells. *Biochim Biophys Acta* 1786: 4–14.
79. Ito H, Gardner-Thorpe J, Zinner MJ, Ashley SW, Whang EE (2003) Inhibition of tyrosine kinase Src suppresses pancreatic cancer invasiveness. *Surgery* 134: 221–226.
80. Levitzki A, Mishani E (2006) Tyrosine kinase inhibitors. *Annu Rev Biochem* 75: 93–109.
81. Hong SM, Park JY, Hruban RH, Goggins M (2011) Molecular signatures of pancreatic cancer. *Arch Pathol Lab Med* 135: 716–727.
82. Tan AC, Jimeno A, Lin SH, Wheelhouse J, Chan F, et al. (2009) Characterizing DNA methylation patterns in pancreatic cancer genome. *Mol Oncol* 3: 425–438.
83. Zhu Y, Zhang JJ, Zhu R, Zhu Y, Liang WB, et al. (2011) The increase in the expression and hypomethylation of MUC4 gene with the progression of pancreatic ductal adenocarcinoma. *Med Oncol* 28 Suppl 1: S175–84.
84. Matsubayashi H, Sato N, Fukushima N, Yeo CJ, Walter KM, et al. (2003) Methylation of cyclin D2 is observed frequently in pancreatic cancer but is also an age-related phenomenon in gastrointestinal tissues. *Clin Cancer Res* 9: 1446–1452.
85. Baylín SB, Jones PA (2011) A decade of exploring the cancer epigenome - biological and translational implications. *Nat Rev Cancer* 11: 726–734.
86. Maier T, Guell M, Serrano L (2009) Correlation of mRNA and protein in complex biological samples. *FEBS Lett* 583: 3966–3973.

87. Cecconi D, Palmieri M, Donadelli M (2011) Proteomics in pancreatic cancer research. *Proteomics* 11: 816–828.
88. Sun C, Rosendahl AH, Ansari D, Andersson R (2011) Proteome-based biomarkers in pancreatic cancer. *World J Gastroenterol* 17: 4845–4852.
89. Bachet JB, Marechal R, Demetter P, Bonnetain F, Cros J, et al. (2013) S100A2 is a predictive biomarker of adjuvant therapy benefit in pancreatic adenocarcinoma. *Eur J Cancer* 49: 2643–2653.
90. Marenholz I, Heizmann CW, Fritz G (2004) S100 proteins in mouse and man: from evolution to function and pathology (including an update of the nomenclature). *Biochem Biophys Res Commun* 322: 1111–1122.
91. Nakata K, Nagai E, Ohuchida K, Hayashi A, Miyasaka Y, et al. (2010) S100P is a novel marker to identify intraductal papillary mucinous neoplasms. *Hum Pathol* 41: 824–831.
92. Dim DC, Jiang F, Qiu Q, Li T, Darwin P, et al. (2011) The usefulness of S100P, mesothelin, fascin, prostate stem cell antigen, and 14-3-3 sigma in diagnosing pancreatic adenocarcinoma in cytological specimens obtained by endoscopic ultrasound guided fine-needle aspiration. *Diagn Cytopathol*
93. Goodell CA, Belisle JA, Gubbels JA, Migneault M, Rancourt C, et al. (2009) Characterization of the tumor marker muc16 (ca125) expressed by murine ovarian tumor cell lines and identification of a panel of cross-reactive monoclonal antibodies. *J Ovarian Res* 2: 8.
94. Kessenbrock K, Plaks V, Werb Z (2010) Matrix metalloproteinases: regulators of the tumor microenvironment. *Cell* 141: 52–67.
95. Roy R, Yang J, Moses MA (2009) Matrix metalloproteinases as novel biomarkers and potential therapeutic targets in human cancer. *J Clin Oncol* 27: 5287–5297.
96. Thiery JP (2002) Epithelial-mesenchymal transitions in tumour progression. *Nat Rev Cancer* 2: 442–454.
97. Hua H, Li M, Luo T, Yin Y, Jiang Y (2011) Matrix metalloproteinases in tumorigenesis: an evolving paradigm. *Cell Mol Life Sci* 68: 3853–3868.
98. Singh P, Srinivasan R, Wig JD (2011) Major molecular markers in pancreatic ductal adenocarcinoma and their roles in screening, diagnosis, prognosis, and treatment. *Pancreas* 40: 644–652.
99. Pryczynicz A, Guzinska-Ustymowicz K, Dymicka-Piekarska V, Czyzewska J, Kemona A (2007) Expression of matrix metalloproteinase 9 in pancreatic ductal carcinoma is associated with tumor metastasis formation. *Folia Histochem Cytobiol* 45: 37–40.
100. Yokoyama M, Ochi K, Ichimura M, Mizushima T, Shinji T, et al. (2002) Matrix metalloproteinase-2 in pancreatic juice for diagnosis of pancreatic cancer. *Pancreas* 24: 344–347.
101. Jones LE, Humphreys MJ, Campbell F, Neoptolemos JP, Boyd MT (2004) Comprehensive analysis of matrix metalloproteinase and tissue inhibitor expression in pancreatic cancer: increased expression of matrix metalloproteinase-7 predicts poor survival. *Clin Cancer Res* 10: 2832–2845.
102. Motrescu ER, Rio MC (2008) Cancer cells, adipocytes and matrix metalloproteinase 11: a vicious tumor progression cycle. *Biol Chem* 389: 1037–1041.
103. Kren L, Goncharuk VN, Krenova Z, Stratil D, Hermanova M, et al. (2006) Expression of matrix metalloproteinases 3, 10 and 11 (stromelysins 1, 2 and 3) and matrix metalloproteinase 7 (matrilysin) by cancer cells in non-small cell lung neoplasms. *Clinicopathologic studies. Cesk Patol* 42: 16–19.
104. Wasenius VM, Hemmer S, Kettunen E, Knuutila S, Franssila K, et al. (2003) Hepatocyte growth factor receptor, matrix metalloproteinase-11, tissue inhibitor of metalloproteinase-1, and fibronectin are up-regulated in papillary thyroid carcinoma: a cDNA and tissue microarray study. *Clin Cancer Res* 9: 68–75.
105. Mellick AS, Blackmore D, Weinstein SR, Griffiths LR (2003) An assessment of MMP and TIMP gene expression in cell lines and stroma - tumour differences in microdissected breast cancer biopsies. *Tumour Biol* 24: 258–270.
106. Bauvois B (2012) New facets of matrix metalloproteinases MMP-2 and MMP-9 as cell surface transducers: outside-in signaling and relationship to tumor progression. *Biochim Biophys Acta* 1825: 29–36.
107. Burrell RA, McGranahan N, Bartek J, Swanton C (2013) The causes and consequences of genetic heterogeneity in cancer evolution. *Nature* 501: 338–345.
108. Campbell PJ, Yachida S, Mudie LJ, Stephens PJ, Pleasance ED, et al. (2010) The patterns and dynamics of genomic instability in metastatic pancreatic cancer. *Nature* 467: 1109–1113.
109. Zahreddine H, Borden KL (2013) Mechanisms and insights into drug resistance in cancer. *Front Pharmacol* 4: 28.
110. Sato N, Fukushima N, Maitra A, Matsubayashi H, Yeo CJ, et al. (2003) Discovery of novel targets for aberrant methylation in pancreatic carcinoma using high-throughput microarrays. *Cancer Res* 63: 3735–3742.
111. Fitzgerald M, Oshiro M, Holtan N, Krager K, Cullen JJ, et al. (2003) Human pancreatic carcinoma cells activate maspin expression through loss of epigenetic control. *Neoplasia* 5: 427–436.
112. Ohike N, Maass N, Mundhenke C, Biallek M, Zhang M, et al. (2003) Clinicopathological significance and molecular regulation of maspin expression in ductal adenocarcinoma of the pancreas. *Cancer Lett* 199: 193–200.
113. Kwon MJ, Kim SH, Jeong HM, Jung HS, Kim SS, et al. (2011) Claudin-4 overexpression is associated with epigenetic derepression in gastric carcinoma. *Lab Invest* 91: 1652–1667.
114. Shiba-Ishii A, Noguchi M (2012) Aberrant stratifin overexpression is regulated by tumor-associated CpG demethylation in lung adenocarcinoma. *Am J Pathol* 180: 1653–1662.
115. Vestergaard EM, Nexø E, Tørring N, Borre M, Orntoft TF, et al. (2010) Promoter hypomethylation and upregulation of trefoil factors in prostate cancer. *Int J Cancer* 127: 1857–1865.
116. Rosty C, Ueki T, Argani P, Jansen M, Yeo CJ, et al. (2002) Overexpression of S100A4 in pancreatic ductal adenocarcinomas is associated with poor differentiation and DNA hypomethylation. *Am J Pathol* 160: 45–50.
117. Nelson HH, Almquist LM, LaRocca JL, Plaza SL, Lambert-Messerlian GM, et al. (2011) The relationship between tumor MSLN methylation and serum mesothelin (SMRP) in mesothelioma. *Epigenetics* 6: 1029–1034.
118. Li Y, Meng G, Guo QN (2008) Changes in genomic imprinting and gene expression associated with transformation in a model of human osteosarcoma. *Exp Mol Pathol* 84: 234–239.
119. Dimberg J, Strom K, Lofgren S, Zar N, Lindh M, et al. (2012) DNA promoter methylation status and protein expression of interleukin-8 in human colorectal adenocarcinomas. *Int J Colorectal Dis* 27: 709–714.
120. Komatsu Y, Waku T, Iwasaki N, Ono W, Yamaguchi C, et al. (2012) Global analysis of DNA methylation in early-stage liver fibrosis. *BMC Med Genomics* 5: 5.